REMARKS

The Office Action mailed on November 2, 2006 has been given careful consideration by applicant. Reconsideration of the application is requested in view of the amendments and comments herein. Claims 1, 5, 6, 11, 15, and 20 have been amended.

The Office Action

Claims 1-2, 4-5, 10-12, 14-15, and 20 are rejected under 35 U.S.C. §103(a) as being unpatentable over Simske (U.S. PG Publication No. 2004/0133560) (hereafter Simske) in view of Henkin et al. (hereafter Henkin) (U.S. PG Publication No. 2002/0107735);

Claims 3, 6-7, 13, and 16-17 are rejected under 35 U.S.C. §103(a) as being unpatentable over Simske and Henkin in further view of Kubota (U.S. Patent No. 6,041,323);

Claims 9 and 19 are rejected under 35 U.S.C. §103(a) as being unpatentable over Simske and Henkin in further view of Drissi et al. (U.S. PG Publication No. 2003/0149686); and

Claims 8 and 18 are rejected under 35 U.S.C. §103(a) as being unpatentable over Simske.

First Obviousness Rejection

The examiner has rejected claims 1-2, 4-5, 10-12, 14-15, and 20 under 35 U.S.C. §103(a) as being unpatentable over Simske (U.S. PG Publication No. 2004/0133560) in view of Henkin et al. (U.S. PG Publication No. 2002/0107. This rejection should be withdrawn for at least the following reasons. Simske and Henkin individually or in combination do not teach or suggest the subject invention as set forth in the subject claims.

As amended, independent claim 1 (and similarly independent claim 11) recites a method for computing a measure of similarity between a first (or input) document and one or more disparate (or search results) documents. A first list of rated keywords extracted from the first document and a list of rated keywords extracted from each of the one or more disparate documents are received. The first list of rated keywords and the list of rated keywords from each of the one or more disparate documents are used to determine whether the first document forms part of the one or more disparate documents using a first

computed percentage indicating what percentage of keyword ratings in the first list also exist in the list of at least one of the one or more disparate documents. A percentage is computed for each of the one or more disparate documents indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents when the first computed percentage indicates that the first document is included in at least one of the one or more disparate documents. The first computed percentage is used to specify the measure of similarity when the computed percentage for at least one of the one or more disparate documents is greater than the first computed percentage. The one or more disparate documents are ranked based on the percentage computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents. Simske and Henkin individually or in combination do not teach or suggest such claimed aspects of the subject invention.

In particular, Simske or Henkin do not teach or suggest comparing a first list of rated keywords extracted from a first document to a list of rated keywords and comparing the received extracted from each of the one or more disparate documents that are received, wherein the one or more disparate documents are ranked based on the percentage computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents. Instead, Simske and Henkin are silent with regard to ranking documents. Moreover, they do not mention ranking based on a percentage computed that indicates what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents.

Additionally, neither Simske nor Henkin mention rating keywords are based at least in part by a relevant weight from their associated document language, as recited in independent claim 11. In contrast, Simske and Henkin contemplate computation of a mean shared weight of extended words. Such computation does not employ rating of keywords part by a relevant weight from their associated document language.

Moreover, Simske does not teach or suggest using keywords to determine whether a first document forms part of a second document using a first computed percentage that

indicates what percentage of keyword ratings in the first list also exist in the second list. Instead, as admitted by the examiner, Simske utilizes shared word weights to compare one document to another. This comparison employs one or more of a mean, a maximum and a minimum shared word weight to determine if one document is similar to one or more disparate documents. Simske, paragraph 54. There is no mention, and Simske does not contemplate using a percentage of keyword ratings in a first list to compare to a percentage of keyword ratings in a second list. The shared word weight employed by Simske is related to the type of word (e.g., noun, verb, adjective, etc.), the text font (e.g., boldface, italics, etc.), layout, etc. Simske, paragraphs 29-33. Simske does not teach or suggest computing a keyword rating percentage of any type. Further, there is no mention of utilizing such keyword rating percentage to indicate a percentage of keyword ratings in one list (e.g. from a first document) to a second list (e.g., from a second document).

Furthermore, Simske does not teach or suggest a second percentage that is computed that indicates what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the second list when the first computed percentage indicates that the first document is included in the second document. Instead, Simske teaches computing a mean shared weight of extended words. The mean shared weight is a sum of all word weight values divided by the number of documents to produce a mean value of all relevant word weights. This is not a percentage value that indicates what percentage of keyword ratings along with neighboring keyword ratings in a first list also exist in a second list when a first computed percentage indicates that a first document is included in a second document, as recited in the subject claims. Instead, it is an average computation that produces a mean value of one or more word weights.

Moreover, Simske does not teach or suggest a first computed percentage used to specify the measure of similarity when the second computed percentage is greater than the first computed percentage. There is no mechanism taught in Simske to determine when a second percentage of keyword ratings in greater than a first percentage. Further, Simske does not teach or suggest a first computed percentage that is used to specify a measure of similarity between documents as discussed above.

Independent claim 20 recites an article of manufacture for computing a measure of similarity between a first (or input) document and a second (or search results) document.

The article of manufacture comprises computer usable media including computer readable instructions embedded therein that causes a computer to perform a method. A first list of rated keywords extracted from the first document and a second list of rated keywords extracted from the second document are received. The first and second lists of rated keywords are used to determine whether the first document forms part of the second document using a first computed percentage indicating what percentage of keyword ratings in the first list also exist in the second list. A second percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the second list is computed when the first computed percentage indicates that the first document is included in the second document. The first computed percentage is used to specify the measure of similarity when the second computed percentage is greater than the first computed percentage. If the first computed percentage does not indicate that the first document is included in the second document, a third percentage is computed using the Jaccard distance measure. If the third computed percentage indicates that the first document is a revision of the second document, a fourth percentage is computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list. The fourth computed percentage is used to specify the measure of similarity except when: (i) the fourth computed percentage is greater than the second computed percentage; (ii) the first list of rated keywords is identified using OCR; (iii) the fourth computed percentage is greater than fifty percent; and (iv) less than twenty percent of the keywords in the first list of keywords are in the second list of keywords. Simske and Henkin individually or in combination do not teach or suggest such claimed aspects of the subject invention.

More particularly, neither Simske nor Henkin teach or suggest if the first computed percentage does not indicate that the first document is included in the second document, a third percentage is computed using the Jaccard distance measure. A Jaccard distance measures dissimilarity between sample sets, is obtained by dividing the difference of the sizes of the union and the intersection of two sets by the size of the union, or, simpler, by subtracting the Jaccard coefficient from 1. See, e.g., http://en.wikipedia.org/wiki/Jaccard_index. Simske and/or Henkin do not teach or suggest a Jaccard distance measurement.

In addition, neither Simske nor Henkin teach or suggest computation of a fourth percentage if the third computed percentage indicates that the first document is a revision of the second document. The fourth percentage indicates what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list. The fourth computed percentage is used to specify the measure of similarity except when: (i) the fourth computed percentage is greater than the second computed percentage; (ii) the first list of rated keywords is identified using OCR; (iii) the fourth computed percentage is greater than fifty percent; and (iv) less than twenty percent of the keywords in the first list of keywords are in the second list of keywords. Instead, there is no mention of computing a fourth percentage if a third computed percentage indicates that the first document is a revision of a second document.

For at least the aforementioned reasons, Simske and Henkin individually or in combination do not teach or suggest the subject invention as recited in independent claims 1, 11, or 20 (or claims 2-10 and 12-19 which respectively depend therefrom). Accordingly, withdrawal of this rejection is respectfully requested.

Second Obviousness Rejection

The examiner has rejected claims 3, 6-7, 13, and 16-17 under 35 U.S.C. §103(a) as being unpatentable over Simske and Henkin in further view of Kubota (U.S. Patent No. 6,041,323). This rejection should be withdrawn for at least the following reasons. Claims 3, 6-7, 13, and 16-17 depend from independent claims 1 and 11 respectively, and the field of invention does not make up for the aforementioned deficiencies of Simske and Henkin regarding comparing a first list of rated keywords extracted from a first document to a list of ⌐rated keywords and comparing the received extracted from each of the one or more disparate documents that are received, wherein the one or more disparate documents are ranked based on the percentage computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents. Thus, for at least the reasons discussed above with respect to claims 1, 11 and 20, the combination of Simske, Henkin and Kubota do not teach or suggest the subject claims. Accordingly, the rejection of these claims should be withdrawn.

Third Obviousness Rejection

The examiner has rejected claims 9 and 19 under 35 U.S.C. §103(a) as being unpatentable over Simske and Henkin in further view of Drissi et al. (U.S. PG Publication No. 2003/0149686). This rejection should be withdrawn for at least the following reasons. Claims 9 and 19 depend from independent claims 1 and 11 respectively, and the field of invention does not make up for the aforementioned deficiencies of Simske and Henkin regarding comparing a first list of rated keywords extracted from a first document to a list of rated keywords and comparing the received extracted from each of the one or more disparate documents that are received, wherein the one or more disparate documents are ranked based on the percentage computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents. Thus, for at least the reasons discussed above with respect to claims 1, 11 and 20, the combination of Simske, Henkin and Drissi do not teach or suggest the subject claims. Accordingly, the rejection of these claims should be withdrawn.

Fourth Obviousness Rejection

The examiner has rejected claims 8 and 18 under 35 U.S.C. §103(a) as being unpatentable over Simske. This rejection should be withdrawn for at least the following reasons. Claims 8 and 18 depend from independent claims 1 and 11 respectively and, as noted above, Simske does not teach or suggest comparing a first list of rated keywords extracted from a first document to a list of rated keywords and comparing the received extracted from each of the one or more disparate documents that are received, wherein the one or more disparate documents are ranked based on the percentage computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list also exist in the list for at least one of the one or more disparate documents. Thus, for at least the reasons discussed above with respect to claims 1, 11 and 20, Simske does not teach or suggest the subject claims. Accordingly, the rejection of these claims should be withdrawn.
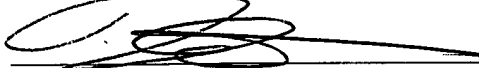
CONCLUSION

For the reasons detailed above, it is submitted that the claims in the subject application are now in condition for allowance. The foregoing comments do not require unnecessary additional search or examination.

In the event the Examiner considers personal contact advantageous to the disposition of this case, he/she is hereby authorized to call Mark Svat, at Telephone Number (216) 861-5582.

Respectfully submitted,

FAY SHARPE LLP

3/2/07
Date

Mark S. Svat, Reg. No. 34,261
Kevin M. Dunn, Reg. No. 52,842
1100 Superior Avenue, Seventh Floor
Cleveland, OH 44114-2579
216-861-5582

CERTIFICATE OF MAILING OR TRANSMISSION	
☒ I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being deposited with the United States Postal Service as First Class Mail, addressed to: Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date indicated below. ☐ transmitted to the USPTO by electronic transmission via EFS-Web on the date indicated below.	
Express Mail Label No.:	Signature: Elaine M. Checovich
Date: 3-2-07	Name: Elaine M. Checovich

N:\XERZ\201374\EMC0005647V001.DOC

14